DOCUMENT RESUME

ED 174 265 IR 007 774

TITLE	The Subject Access Problem--Opportunities for Solution: A Workshop.
INSTITUTION	Committee for the Coordination of National Bibliographic Control, Washington, D.C.
SPONS AGENCY	Council on Library Resources, Inc., Washington, D.C.; National Commission on Libraries and Information Science, Washington, D. C.; National Endowment for the Humanities (NFAH), Washington, D.C.; National Science Foundation, Washington, D.C.
PUB DATE	Oct 78
NOTE	81p.; Workshop given by The Committee for the Coordination of National Bibliographic Control (Springfield, Virginia, October 18-20, 1978)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Abstracting; *Change Strategies; Cooperation; Indexing; Information Needs; *Information Retrieval; *Problems; Standards
IDENTIFIERS	Authority Files; *Subject Access

ABSTRACT

This report presents the recommendations of 35 workshops participants, drawn from a variety of backgrounds and information environments, which was organized to: (1) highlight the role of subject access with the emphasis on the needs of the information seeker; (2) compare techniques for subject access now being used in the library and the abstracting and indexing (A&I) communities, including applicable research efforts; (3) emphasize the role and nature of authority files in subject access control and facilitation; and (4) seek opportunities to improve information preparation for subject access through better functional interconnection and integration across all communities, and to promote the use and enhance techniques for subject access to classes of publications not now adequately covered. The report includes an executive summary, background information on the subject approach to accessing information, a description of the workshop design, 20 definitions of "subject access," a summary of workshop proceedings, draft recommendations from the discussion groups, and a final synthesis with action recommendations. Appendices provide the keynote address by F. Wilfred Lancaster, (Subject Access in Perspective), an address by Margaret T. Fischer (Subject Access: Challenges for the Future), and lists of participants and observers. (Author/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174265

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE SUBJECT ACCESS PROBLEM--OPPORTUNITIES FOR SOLUTION

A Workshop

Sponsored by

The Committee for the Coordination of National Bibliographic Control
Springfield, Virginia
18-20 October 1978

IR001774

The materials incorporated in this report were developed with the financial support of the following:

Council on Library Resources

National Commission on Libraries and Information Science

National Endowment for the Humanities

National Science Foundation

However, the opinions, findings, conclusions, and recomendations expressed herein are those of the authors and do not necessarily reflect the views of the funding agencies identified above.

Members of the Planning Committee

Chairman - Ronald L. Wigington, Director, Research & Development
Chemical Abstracts Service

Toni Carbo Bearman, Executive Director
National Federation of Abstracting and Indexing Services

Lucille Gordon, Manager, Library Marketing
McGraw-Hill Book Company

Paul B. Lagueux, Executive Secretary
Committee for the Coordination of National Bibliographic Control

Jay K. Lucker, Director of Libraries
Massachusetts Institute of Technology

Margaret K. Park, Director, Information Systems
Office of Computing Activities, University of Georgia

Mary K. Pietris, Chief, Subject Cataloging Division
Library of Congress

Table of Contents

	Page
Executive Summary	V
Introduction	1
Background	2
Workshop Design	7
Definition - What Is Subject Access?	10
Summary of Workshop Proceedings	17
Draft Recommendations of Discussion Groups	30
Synthesis and Action Recommendations	37

Appendices

A - Subject Access in Perspective
F. Wilfrid Lancaster

B - Subject Access: Challenges for the Future
Margaret T. Fischer

C - Participants

EXECUTIVE SUMMARY

A workshop entitled "The Subject Access Problem -- Opportunities for Solution" was held in Springfield, Virginia, October 18-20, 1978. The workshop was sponsored by the Committee for the Coordination of National Bibliographic Control which is supported by the National Science Foundation, the National Commission on Libraries and Information Science, and the Council on Library Resources. Funds for the conduct of the workshop were provided by the Committee with additional financial support provided by the National Endowment for the Humanities and the Council on Library Resources.

The need to improve the ability of the user to find needed information through subject access was the basis for the workshop. The Committee organized the workshop to:

- highlight the role of subject access with the emphasis on the needs of the information seeker,
- compare techniques for subject access now being used in the library and the abstracting and indexing (A&I) communities including applicable research efforts,
- emphasize the role and nature of authority files in subject access control and facilitation,
- seek opportunities to (1) improve information preparation for subject access through better functional interconnection and integration across all communities and (2) promote the use and enhance techniques for subject access to classes of publications not now adequately covered.

The planning committee assumed the responsibility for synthesizing the recommendations from the deliberations of the workshop. While endeavoring to express the themes which were most prevalent, the planning committee does not wish to imply that the highlights which are presented in the chapter "Synthesis

and Action Recommendations" are a formal consensus of the participants. Major themes which stood out strongly during the workshop were:

- diversity in subject access needs and approaches is real, is good, and will continue,
- subject access mechanisms should be designed for end users and not exclusively for information professionals,
- "post-coordinate" natural language subject approaches for retrieval are likely to become more emphasized than traditional pre-coordinate subject heading systems to enable easier adaptation to changing interests and needs,
- groups having common interests for serving specific subject access needs should unite for action; these groups must include all relevant aspects of libraries, publishers, abstracting and indexing organizations, and other information agencies,
- some areas (e.g., the humanities), that are now deficient in subject access can learn from the more well developed areas but must adopt techniques selectively and knowingly,
- universal standardization should not itself be a goal; in many situations interlinking of the various communities of interest (e.g., with respect to vocabularies) is more practical and acceptable,
- a single national plan for subject access with centralized funding is not feasible in the pluralistic society and economy of the United States.

The planning committee recommended certain specific actions that should be taken and pointed out that each recommendation requires organizing, funding, and a set of individuals to carry out the work. Those who are capable of providing the necessary initiative and support in furtherance of the overall objectives are not identified in this paper, however, the specifics of this

paper have been brought to their attention. The summary of action recommendations follows:

1. Interlinking of library- and A&I- produced subject access tools
   a. Locate action agents in each of the library and A&I communities who are willing to collaborate in the development of demonstration projects in limited subject areas of the sciences that would focus on how to solve subject access problems in an integrated environment.
   b. In a similar manner, carry out a demonstration project on the linkage between formal catalogs and access tools to manuscripts and other forms of information in a specific area of interest to the humanities.
   c. Additional efforts are needed to broaden awareness and understanding of the various kinds of subject access techniques and tools through:
      1) joint sessions and/or special conferences sponsored by professional associations;
      2) publications, including up-to-date textbooks and curriculum materials, that emphasize the integration of available resources.

2. A "system view" of future subject-access mechanisms
   a. Conduct a research project on the design characteristics of a large dynamic file system based on file usage data from the users and from collection change and growth to examine
      1) use patterns and user behavior,
      2) collection purging criteria and phenomena,
      3) user or use-created cross-reference systems.

3. Gaps in subject access tools for the general population
   a. Make and disseminate an analysis of the availability and suitability of subject access approaches and tools for users of school and public libraries.

4. Subject access for the humanities
   a. Conduct a study to determine
      1) the nature of humanities literature,
      2) how humanists use information,
      3) the design of tools appropriate to the characteristics of the literature.
   b. Develop the capability to review, explore, and evaluate techniques in other fields in order to best apply relevant techniques.

5. Subject access to monographs
   a. Encourage and support projects that offer innovative and practical solutions to the problem of in-depth subject access to monographs.
   b. Conduct an experimental research project to study the generation of index material from monographs in machine-readable form as a by-product of computer controlled composition.

6. Subject access authority files
   a. Include the means for linking library and A&I produced subject-access tools by including participation of both communities in the development of a rational authority system for the Bibliographic Service Development Program.
   b. As a related effort, assemble a group of experts to study the feasibility of defining codes of practice, where necessary, to guide the construction and use of each subject-access subsystem of the overall authority system.

INTRODUCTION

The Committee for the Coordination of National Bibliographic Control (CCNBC) was established to pursue specific, common problems of concern to all sectors of the information community. In its efforts to enhance bibliographic control -- those principles and processes by which bibliographic items are identified to the level required for the management of and intellectual access to information of all types -- the Committee has relied upon the cooperation those who publish, process, store, and disseminate information in all of its forms.

The Committee has done its work through a variety of means, including chartering specific studies, reviewing various projects and activities, advising its sponsors, and organizing workshops. In this case, recognizing that subject access to information is a diverse set of interests and activities and that it is timely to focus more attention on the status, problems, and developments relating to subject access, the Committee decided to hold a workshop to explore these matters and appointed a planning committee to organize and conduct it.

Participants were invited from a wide variety of backgrounds and environments; a complete list is given in an appendix to this report. Support for conducting the workshop was obtained from the National Endowment for the Humanities, the Council on Library Resources, and from Committee funds.

This report represents the completion of the CCNBC project to focus attention on subject access. However, this report alone is not sufficient to accelerate the actual solution of subject access problems facing information users. Organizations that have the means to cause things to happen through their funding or their operating plans must pick up these themes and bring about real and beneficial changes.

BACKGROUND

Of the several modes of retrieving information, the subject approach holds the greatest and most dynamic potential for serving information seekers, yet it is the least understood, the least standardized, the least developed, and, often, the most costly process. Subject-access systems may be characterized as customized, oriented by discipline, based on historical precedents, and, usually, constrained by collection size and financial considerations. The scope of "subject access" encompasses the concerns of libraries and their subject catalogs, of abstracting and indexing services and their publications and services, of archivists and their inventories, and of publishers and their catalogs. Practices in these fields have developed independently, with the result that there is little commonality or compatibility.

One of the several uses of the catalog in libraries is its employment as a tool for subject access -- to locate sources of information about a particular subject. The primary emphasis of card catalogs is on monographs. On the other hand, access to the subject content of articles within serials and the report literature has been emphasized by abstracting and indexing publications and services. In both cases other forms of publications are represented, but in each case the nature of the document determines the approaches and techniques that are utilized.

Although subject cataloging is a time-honored speciality of library science and although subject indexing is the foundation of abstracting and indexing services, there is insufficient collaboration and integration of efforts toward subject access among the publishing, abstracting and indexing, and library communities. And yet, from the information seeker's point of view, techniques and mechanisms of subject access are at the heart of information retrieval. The need to improve the ability

of the user to find needed information through subject access was the basis for this workshop.

The topic is timely: card catalogs are being phased out and their function must be replaced; book catalogs, COM catalogs, and online access are emerging; and the proliferation of systems for subject access is becoming more obvious to users attempting to approach a variety of dissimilar data bases.

In view of these problems, the Commitee for the Coordination of National Bibliographic Control organized a workshop to:

- highlight the role of subject access in library and information systems, with emphasis on the needs of the information seeker;

- assess the nature and extent to which there is a subject-access problem and identify related unsolved problems of the field;

- compare the techniques for subject access now being used, including the related research emphasis and results in the library and the abstracting and indexing communities, and seek to improve communications between the two communities with regard to subject-access research efforts;

- emphasize the specific role and nature of authority files in subject-access control and facilitation;

- seek opportunities to (1) improve information preparation for subject access to all classes of publications through better functional interconnection and integration across publishing, abstracting and indexing, and library operations and (2) promote the use and enhancement of techniques for increased subject access to classes of publications

not now adequately covered (e.g., monographs and audiovisuals).

As a result of the traditions of library cataloging and the current impracticality of doing anything more comprehensive, conventional subject cataloging is conservative in the number of access points and their uses, employing the fewest headings needed to cover the focus of a work without providing headings for more specific topics included in the work. While most current cataloging systems are capable theoretically of specificity, various practical considerations often limit the provision of highly specific headings unless a work itself deals only with a specific topic. In-depth cataloging is lacking for general works. This is exemplified in serials, where libraries have largely confined their cataloging efforts to the title level and provide no basis for detailed subject access.

Meanwhile, abstracting and indexing services and information dissemination services are providing subject access to various classes of publications. Some are aimed at maintaining consistency with library subject systems, for example the H. W. Wilson Company's Humanities Index, Social Science Index, Applied Science and Technology Index, etc. Others have evolved into independent, detailed subject-access tools, such as the subject indexes of the abstracting and indexing services, and have interface only at the title (cataloging) level. Yet others have developed a document delivery function, for example, the Educational Resources Information Center (ERIC) and its companion, the ERIC Document Reproduction Service (EDRS). The document delivery function has recently received considerable attention, including the interest of the National Commission on Libraries and Information Science (NCLIS), the National Technical Information Service (NTIS) (although its Journal Article Copy Service [JACS] was recently discontinued), the National Commission on New Technological Uses of Copyrighted Works (CONTU), and others. Another difficult matter to be considered is that libraries and indexing and abstracting services have

avoided identifying content for particular audiences. For example, the treatment of a subject will be much different depending upon whether it is aimed at scholars, the general adult population, or elementary school children.

Two examples of integration of the library with computer-based search access functions into a single organization are the National Library of Medicine and its MEDLARS and the National Library of Canada and its CANOLE. However, even in these cases the approach is based on common management of two kinds of activity and does not represent a true, functional integration.

A major factor in traditional library practice that affects subject access as well as collection management is the growing difficulty in constructing and maintaining large and comprehensive card catalogs. This is reflected in discussions about and, in some cases, plans for replacing card catalogs with, for example, computer-based access systems. Some totally computer-based systems for library management and information retrieval, such as BALLOTS, have included elements of computerized text search functions in their original design.

With the increasing computerization of the information needed for management of and access to information collections, there is a concurrent requirement to bring these two functions into harmony. There is also an opportunity to introduce new methods. It is essential that collective expertise be brought to bear on solving these problems.

The participants in this workshop included subject-access users from various disciplines as well as providers of information -- reference librarians, information specialists, systems designers, and library and information systems managers. The workshop concentrated on significant technical issues and approaches as well as on system design and operating management concerns.

A list of questions was prepared and used to stimulate the discussions:

- How can we eliminate the expensive and time-consuming duplication of intellectual processes in creating subject access to bibliographic entities?

- What are the "elementary particles" of the subject-access universe: subject headings, index terms, key words, descriptions, identifiers, links, roles, classes, text, codes, etc.? What are their differences and similarities?

- What is the impact, or potential impact, on the input process (including cataloging, indexing and abstracting) relative to text searching capabilities in information retrieval systems?

- What practices for indexing that have been developed for specialized use in various disciplines should be more widely adopted for general use?

- Can subject access be provided as effectively for monographs as for items like journal articles and technical reports dealing with more specific subject areas?

- How can libraries effectively provide users with subject access to the collections through data base search services?

- What is the role of present standards that are applicable to subject access? How effective are they? Are new standards needed?

- As files get larger, must subject access necessarily become more complex and expensive? Should files be partitioned -- for example, by date, document type, gross categories, etc., -- as they get larger?

WORKSHOP DESIGN

The workshop was limited to thirty-five invited participants, plus seven observers. It consisted of six sessions beginning with an evening session and followed by two full days. A broad mix of participants was invited in order to obtain a full range of points of view and experience.

The workshop used an analytic approach. A set of questions was posed: What is the subject-access problem? How is subject access handled? What are the related questions that lead to problems? The special characteristics of each problem were considered, with full attention given to differences arising from such things as discipline or subject area, type of material or medium, etc. The second part of the analytic approach was to discuss the alleviation of some of the problems that had been identified in the first half of the program and then to attempt to identify prospects for working on them cooperatively.

One of the objectives of the introductory evening session with the keynote speaker was to place "Subject Access in Perspective" as one of many ways of accessing information. Other objectives were to (1) start the thinking that led to answers to such questions as why, when, and how subject access is used and for what purpose; (2) characterize the nature of subject access as distinguished from other ways of finding information in collections; and (3) emphasize the fact that subject access is lagging behind other means of access in terms of technique, degree of automation, and other qualities needed to provide good and complete service to information seekers.

The first full day of the workshop started with a session on "Subject Access: Where are we now, and why haven't we gotten further?" Each member of a panel representing a mix of backgrounds stated briefly a point of view that was related to the theme. General discussion by the workshop participants followed.

Following that, in the afternoon, the workshop discussed "Current Research and Development in Subject Access." During this session current activities that may contribute techniques and experience toward solutions for the future were reviewed and compared. Ideas and results from the diverse backgrounds and knowledge of the participants were shared to establish a broader awareness of work now being done in specialized areas. Existing work on intercommunity interfaces, limits of technology in achieving various solutions, and related cost factors were included in this session.

The evening of the first full day featured "Subject Access: Challenges for the Future." The purpose of the formal presentation was to stimulate the participants' thinking in order to begin the transition from the problem identification and current activity nature of the earlier sessions to analysis and problem solving.

The morning of the second day was devoted to "Subject Access: Key Problem Areas and Issues," in which the discussion was based upon a list of problem areas, issues, and questions developed from the earlier sessions.

An extended noon period provided for lunch in small groups to which participants were assigned. Informal discussions generated recommendations to be considered by the workshop participants in the afternoon. A leader and a recorder were designated for each group to organize the discussions and reporting.

The final session involved all participants in the review and discussion of the suggested recommendations.

The planning committee worked from a summarized transcript prepared by a professional conference recorder to produce a draft of the workshop proceedings. This was circulated to participants for comment and critique. From this draft of the workshop proceedings and the participants' comments and critiques, the planning committee constructed the "Synthesis and Action Recommendations." The report is now being provided to the

sponsors who supported the workshop and to other interested institutions and associations for their consideration or actions.

Definition - WHAT IS SUBJECT ACCESS?

While the term "subject access" was used extensively in the planning and in discussions with the prospective participants before the workshop, it was not realistic to assume that everybody meant the same thing by it. Thus, each participant was invited to furnish a definition before the workshop convened by completing the phrase "By subject access is meant..." The planning committee, independently, constructed its own.

These definitions reproduced below, without attribution, provide a snapshot of the frame of reference of the workshop participants as they assembled.

The planning committtee definition was: "By subject access is meant the use of words, phrases, or symbols to represent the intellectual content of recorded knowledge for purposes of organization and retrieval."

As expressed by the participants, by "subject access" is meant...

1. ...the ability to retrieve information or documents, etc., from their theme or unit concept.

2. ...in the broadest sense, gaining access to a desired document by means of a subject heading. More specifically, as a producer of abstracting and indexing publications, it means locating a document of interest by looking it up in the subject index of a bibliography or online service.

I hope that the consideration of subject access will include a comparative evaluation of current techniques of subject access, as well as a comparative evaluation of different types of authority files. I would further hope that the problem of subject access will be discussed in the context of current and developing technology, especially the computer.

3. ...access to a document by means of its actual content as contrasted to access by means of its title, author names, and routine identifying citations. Such subject content is usually referenced within the document title but often in a very narrow sense. Subject access provides many additional routes to the document not available via the title alone. In today's environment of multidiscipline and mission-oriented interests, this increased subject access is most important. Much important information is buried within documents and is not retrievable by way of routine citation searching.

Subject access has another important objective. It places the content of a given document in a relationship to other documents with the same or similar content. A given subject serves as a central collecting point for all documents of similar content. Since individual subjects gleaned from the document content are more specific than the titles, the resulting access points are more effective for searching. Subject access, unlike a formal title, is amenable to vocabulary control. Such control results in more ordered, more efficient, and less costly searching.

4. ...access to topical information/material by means of descriptors in an index or catalog. Access to topical information/material by means of numerical or letter/number notations (e.g., Dewey Decimal Classification).

5. ...the means by which any given collection of information may be searched by subject, topic, theme, key word(s), or element within a bibliographic entry other than the "standard" ones (which I take to mean, e.g., author's name, title of work, imprint). For the "Cumulative Shakespeare Bibliography" we are preparing under a grant from the NEH, our electronic data-processing system permits access to data contained in all the recognized elements of standard bibliographic entries organized within a comprehensive taxonomy, but also, by employing a

thesaurus-term "seeding" we have the capability to search the file by subject and theme irrespective of the taxonomy.

6. ...that a range of intellectual approaches must be used, in a disciplined manner that are keyed to different levels of arrangement of the material. Serial arrangement and collective description (as distinct from item description) are the main characteristics. Content analysis of record series for their subject matter is only one form of intellectual access to subject matter, and is the most expensive. One of the chief modes, and the least expensive, capitilizes on the user's prior association of proper names with activity/subject matter. A high degree of proper name control correspondingly provides a high degree of subject access, which is independent of content analysis. Entries for place names and names of political subdivisions also provide modes of subject access for those (a small minority) using this route.

7. ...a means of identifying and retrieving a bibliographic entity by querying a data file on the basis of discretely and systematically codified subjects.

8. ...providing effective yet economical access to a heterogeneous mixture of information sources within a well-defined or hierarchical research design.

Subject-access systems unless of the highest priority (such as those necessary for National Defense and Medical Diagnosis), should not provide instant answers to unstructured questions. Subject-access systems should be designed to lead researchers in the direction they want to go through a series of increasingly better defined and more specialized searches for sources to consult. Subject access ultimately means the researcher working with a manageable amount of material, not providing the researcher with answers to questions posed.

9. ...the ability of a person to obtain facts, materials, information, etc., regardless of their age, geographic location, economic status, or social status. It means the access through the traditional means, i.e., a library, and through nontraditional means, community service centers, computer data bases, and other people. Accessing this information for people should be free of the prejudices of those assisting in the accessing, and free of cost to all. These materials and information may be available in a variety of formats also, including book, microform, recordings, video cassette, etc. The responsibility of those locating the information is to search the broadest possible range of formats and sources. This definition of subject access may be very broad - not limited to our traditional use of the card catalog - as I feel that the location of information should go beyond the four walls of a library.

10. ...the operations necessary for the retrieval of recorded data or information from the media on which it is stored by the use of subject concept labels that have been assigned to the data or informaton for purposes of identification or which were inherent in the data or information. The term "subject access" should also include, to the degree that it is necessary for the purposes of the information seeker, timeliness, low cost, relevance, comprehensiveness of recall, ease of use of the system, precision, accuracy, unambiguousness, and provision of a useable replica of the data or information.

11. ...the approaches to subject matter in a collection. The approach may be systematic (as in the classified arrangement of books on a library shelf) or the approach may be topical (as in the subject headings in the card catalog) or the approach may be "natural or free" (as in the title words or words in an abstract or subject description if they are for matching with query words).

All of the above approaches may be considered the result of subject cataloging and/or the features of the retrieval system(s)

available to the searcher trying to find the subject matter in the collection.

Shera and Egan in 1956 said the objects of subject cataloging should be:

a. to provide access by subject to all relevant material;

b. to provide subject access to materials through all suitable principles of subject organization, e.g., matter, process, applications, etc;

c. to bring together references to materials which treat substantially the same subject regardless of disparities in terminology, disparities which may have resulted from national differences, differences among groups of subject specialists, and/or from the changing nature of the concepts within the discipline itself;

d. to show affiliations among subject fields, affiliations which may depend upon similarities of matter studied, or of point of view, or upon use or application of knowledge;

e. to provide entry to any subject field at any level of analysis, from the most general to the most specific;

f. to provide entry through any vocabulary common to any considerable group of users, specialized or lay,

g. to provide a formal description of the subject content of any bibliographic unit in the most precise, or specific, terms possible, whether the description be in the form of a word or brief phrase or in the form of a class number or symbol;

h. to provide means for the user to make selection from among all items in any particular category, according to any chosen set of criteria such as most thorough, most recent, most elementary, etc.

12. ...being able to find out about pink elephants, magenta pachyderms or light red mammoths. It is being able to find information, or references to information, by topic, at varying levels of specificity no matter in what form, i.e., book,

article, conference paper, patent, report, tables, drawings, photographs, bibliographies, or in parts of any of the above.

13. ...a way of coding the content of documents in a collection, by creating a subject vocabulary (alphabetic, numeric, or other), which can then be used to identify documents relevant to a particular search need.

14. ...THE ABILITY OF USERS (such as researchers, library patrons, students, librarians, government officials and information seekers in general)

OF A DATA BASE (such as a card catalog, an online information file, a periodicals index, a publisher's catalog, etc.)

TO LOCATE REFERENCES (such as cataloging information, bibliographic citations, legislative information, etc.)

TO NEEDED INFORMATION BY MEANS OF SUBJECT CONTENT DESIGNATORS as indexing terms, subject headings, descriptors, keywords, etc.).

15. ...an approach to organization of and subsequent retrieval from a body of public knowledge represented by documents, on the basis of their contents and potential for conveying information. Since by public knowledge is meant a rational consensus of ideas and information, subject access is predicated on the existence of a consensus and limited to that consensus.

16. ...a method of retrieving information from a document collection such as a card catalog or machine-readable data base, by addressing the topics which are discussed in the documents or members of the file and which are usually represented as natural language terms or added index terms.

17. ...the search strategy or tactics involved in retrieving the manually, semiautomatically, or automatically indexed content of documents.

18. ...when considered in practical terms, a connection with the online retrieval systems that will prevail in the future. In these circumstances, subject access must then be provided with respect to a wide variety of different criteria, involving large sets of free-form vocabularies. The questions connected with subject access include the handling of information items described by diverse, nonstandardized information identifiers, the handling of incompletely or ambiguously identified information items, the identification of incompletely specified information requests, the matching of information requests with information items.

19. ...identification of relevant records based on subject searching, particularly for concepts that are not easily described (e.g., "street people," "white flight," "honesty in government"). I want to be able to find publications on how to install various kinds of brick walkways in my back yard.

20. ...the means by which people who seek information use words -- a word or word group -- to find what they seek. These words, taken from language, may be controlled or uncontrolled in varying degrees of organization according to a wide variety of schemata, schedules, and classifications.

SUMMARY OF WORKSHOP PROCEEDINGS

The Subject Access Workshop was designed to approach the topic of subject access from a very broad perspective, addressing the range of subject-access systems and, most important, including among the workshop participants representatives from many of the groups involved in subject access. Subject access was viewed by the workshop attendees as a series of filters providing access to information at different levels. Participants in the subject-access system, all involved with this stepwise processing of information, include publishers, abstracting and indexing services, information dissemination centers, libraries, and, ultimately, the end-users. All were represented at the workshop, as were a range of disciplines including the sciences and the humanities. Educators, researchers, and funders in the subject-access area also particpated.

As indicated in the workshop objectives, the emphasis was on identification of the contemporary subject-access issues and related recommendations for action. At the workshop, several types of activities served as background to this major task, including sessions on the current situation and ongoing research and development areas. Addresses on "Subject Access in Perspective" (F. Wilfrid Lancaster) and "Challenges for the Future" (Margaret T. Fischer) are included as appendices to this report. Each activity contributed to a more thorough understanding of the subject-access problem by the diverse group of participants in attendance.

The definitions of subject access provided by the workshop attendees prior to the meeting (and reproduced in the previous section of this report) gave an indication of the diversity of the viewpoints that would echo throughout the workshop.

Definitions varied for the types of primary materials and subject-access systems referred to, in the techniques mentioned as appropriate, in the purposes of subject access, and in the degree of emphasis placed on the materials accessed versus the user.

Professor Lancaster's presentation opened the workshop and addressed both the present and the future of subject access. He mentioned many of the topics that would emerge from the workshop as key issues. Major developments since World War II were noted: the development of post-coordinate indexing techniques, the implementation of such methods through increasingly more sophisticated means, and the growth in machine-readable data-base activity and in associated telecommunications activity. At this point, the workshop acknowledged that subject-access systems are complex, the success or failure of which are governed by a number of factors mostly outside the control of the information center providing searches. The complexity of subject-access systems also suggests the need for multiple approaches to system design and the difficulties in evaluating performance. Current gaps in our information primarily involve the user-system interface, including user search behavior.

Speculating on the future, Professor Lancaster noted that subject-access systems will be operating in a changing environment in which information is acquired on demand in a pay-as-you-go process. Use will switch from a delegated or mediated mode to direct user-system interaction. Search systems will utilize both the conventional and newer approaches, including some which are language independent. An unlimited variety of data bases will be available to the user, and the system will assist in the selection of both data bases and references within a data base. The language of the systems is likely to be hybrid -- natural language input and a post-controlled vocabulary at output. This picture of the future stimulated workshop participants to consider trends of present developments and to evaluate them in terms of the overall goals of subject-access systems.

The session on "Subject Access: Where are we now, and why haven't we gotten further?" included presentations by four workshop participants representing libraries, abstracting and indexing services, information dissemination centers, and publishers. Each discussed the current state of affairs relative to subject access in his segment of the information community, goals and objectives, and some of the major problems encountered. The individual presentations also brought out some of the interactions among the different segments of the community represented.

Mary Kay Pietris, from the Library of Congress, addressed the topic of development and use of LC subject headings. They are, she pointed out, both widely used and widely criticized. Criticisms relate mainly to their use of outdated and/or inappropriate terminology, to various inconsistencies, and to time delays in production. Barriers to resolving these problems include the lack of a code, or theoretical basis, for the practices followed and the difficulty of modifying large files of back records both at LC and in other libraries to make them consistent with newly adopted practices.

Russell J. Rowlett, Jr., from the Chemical Abstracts Service, described the goal of subject indexing by the abstracting and indexing services as the provision of access to the contents of documents, one step in a series of filtrations of information available to the potential user. Subject-access issues in the abstracting and indexing community include authority control, cooperative terminology control, and controlled versus uncontrolled search vocabularies.

Kay Durkin of Bibliographic Retrieval Services pointed out that considerable work has been done on a file by the time it reaches the information dissemination center and that this constrains further activities. Other constraints are imposed by the search system used. Information dissemination centers are mainly responsible for formatting bibliographic tapes for searching and for developing appropriate search systems. The

diversity of file formats is a problem in loading the files and is also a problem in the education of users by the information dissemination centers.

Lucille Gordon of McGraw-Hill discussed the role of the commercial publisher in providing subject access. She pointed out that the limited subject-access activities carried out -- primarily back of the book indexing and assignment of broad subject categories -- are seen as sufficient by most publishers. An incentive to change might be provided by the marketplace, but libraries are not a highly visible part of that marketplace, and there must be some business advantage to warrant doing the extra work.

Several commonalities were identified by workshop participants in the presentations given. In viewing the overall flow of materials from author to user, with all the attendant subject-access activities, a trend toward centralization was seen by some (but not by others), particularly in the form of a national bibliographic data base. The need for awareness of other system participants, and a willingness to cooperate with them to improve subject access, was also a common thread of the presentations. The existence of, and the need to build upon, a diversity of subject-access systems was recognized and the key issue of linkages between systems was raised by several participants.

A second background session on "Current Research and Development in Subject Access" included a sampling of a number of current research efforts in the subject-access area. Topics covered included the Subject-Access Project recently completed by Syracuse University (Pauline Atherton), the Preserved Context Index System (PRECIS) (David Batty), and subject-access activities at OCLC, Inc., (formerly Ohio College Library Center) (Ann Ekstrom). Shorter presentations were also given on subject-access activities at the Chemical Abstracts Service (Russell Rowlett), subject access in the humanities (Richard Lineback), development of the Congressional Research Service thesaurus

(Frederick Rosenthal), and automatic indexing (Everett Brenner). Generally these presentations gave an indication of subject-access techniques currently being tested or in use. Again, the picture was of considerable diversity, ranging, for example, from the highly structured PRECIS to several examples of "natural language" approaches. A user orientation was strong in most of the presentations, and the importance of system interfaces and the need for building upon existing systems were also emphasized.

The transition for the workshop group from concentration on the current status of subject access to a future-oriented discussion of issues and recommendations was provided by Margaret Fischer in her address "Challenges for the Future." She reminded the participants that the goal of subject access is to provide information on all levels, in all depths, to all kinds of people. She emphasized the different types of primary materials and different types of user needs and the opportunities available to develop responsive systems. A series of challenges was set out: to work together to provide subject access, to cover all forms and types of information, to respond to the marketplace, to use technology in support of subject-access systems, to build associative systems, and others. Finally, she expressed the basic challenge as that of maintaining knowledge as a viable unity through sifting, reviewing and synthesizing, moving towards H. G. Wells's concept of a dynamic world encyclopedia.

The next workshop session dealt directly with subject-access issues, with participants invited to put forth key problem areas and concerns which they had indentified. In this way, a list of twenty-five issues was created and discussed. Most of these related to the major themes evident in earlier workshop discussions.

"Diversity" was the conference key word, emerging over and over again as a characteristic associated with every aspect of subject access. Diversity was seen in the materials processed, in organizations doing the processing, in techniques for subject access, and, perhaps most important, in users of subject-access

systems. Generally, diversity was seen as positive, as the only means of dealing with the given variations in materials and users. Centralization and standardization were discussed, and were found in most cases to be less desirable than the maintenance of diverse systems with bridges or interfaces between them. In the area of techniques, post- rather than pre-coordinate systems were seen as preferable, with hybrid systems a possibility. Several bases for new techniques were mentioned. Throughout the discussions, the need to focus on the users -- to understand their needs and to design systems in response to them -- was emphasized.

The issues identified by workshop participants fell generally into three categories: those associated with the subject-access systems themselves, particularly coverage, vocabularies, and search techniques; those associated with the system environment in some way; and those involving the interfaces between different subject-access systems and between users and systems. They are listed and briefly discussed below.

Issues Relating to Subject-Access Systems

- Diversity of subject-access vocabularies
- Little subject access for some types of publications and for some subject fields
- Outdated subject-access systems
- The extra effort required for expanded recording of subject-access entries
- The need to supplement subject access with quality indicators
- The need for techniques for purging collections

. Pre- versus post-organization of data collections

. Time dependencies of subject-access vocabularies

These issues covered a wide scope, addressing both general concerns and specific system aspects related to vocabularies used, search techniques, and other system characteristics.

Two general problem areas were mentioned: the lack of subject-access systems for some types of publications and for some subject fields, and the outdated nature of some subject-access systems. Differences in coverage for books and journal articles, the inadequacy of subject-access systems in many areas of the humanities, and the limited initial efforts in subject-access systems associated with audiovisual materials and nonbibliographic data bases are examples of the first problem area. The need for appropriate systems for all categories of users, including the general public as well as researchers and educators, was also mentioned. The point was strongly made by a participant that the information needs of poor people, minorities, women, disabled persons, seniors, and prisoners are not adequately met by existing means and that subject-access approaches suitable for these audiences were different than those usually considered by most library and information service institutions. In connection with the question of outdated systems, the workshop participants recognized the related problem of the difficulty of modifying or replacing an already existing system, but saw a need for up-to-date systems which have maximum responsiveness to user needs.

The diversity of system vocabularies may be either a problem or a strength, but certainly it is a fact of life to be dealt with. Here, as elsewhere, the need for bridging techniques between vocabularies was noted. Historical developments alternating between natural language searching and controlled vocabularies were suggested. Another vocabulary problem that was identified concerned the time element; basically, whether the

vocabulary used should be associated with the time of generation of the primary material or the time of its use. And how are linkages between the two frameworks established? This problem has one solution in manual files and possibly another in the newer automated files.

Several issues revolve around the development of user-responsive systems, and three specific techniques were discussed. One issue mentioned throughout the workshop was pre- versus post-coordinate systems, with the advantages and disadvantages of each in different circumstances addressed. A suggestion for supplementing current subject-access systems with quality or nonsubject indicators was made in order to expand the access paths available to users. Appropriate nonsubject indicators should also be included and the ability to search on them provided. Another element of subject-access systems which could provide better services to users would involve maintenance of records of system use. This concept involves both storing of successful search requests for others' use and the use of performance data in system modification and development.

Also, techniques for purging or partitioning files were discussed. Though this issue relates to user considerations, it is more concerned with the technologies utilized by systems and associated cost considerations.

Issues Relating to Subject-Access Environment

- Lack of R&D support for subject-access methods and systems
- Environment needed for initiating change
- Impact of technology on subject-access policies and practices, and user characteristics
- Cost effectiveness of subject-access processes

- Institutional responsibilities and relationships
- Diversity of topic coverage
- Special characteristics of the humanities literature
- Need for a relationship of subject-access to numeric data collections

A number of distinct issues can be identified when one considers the environment in which subject-access systems function. This environment includes materials on the input side and users on the output side, and involves institutional participants including publishers, abstracting and indexing services, information dissemination centers, and libraries. The role of researchers and funders must also be taken into account.

Some of the questions that arise in the evaluation of any large system relate to participant interactions, system costs, and system performance. In the subject-access field, the issue of institutional responsibilities and relationships is a particularly complex one, since the subject-access chain is long and involves participants with widely differing goals and incentives. Analysis of all the subject-access systems for books and monographs suggests a number of structural shifts which have occurred and others which appear likely in the future. Awareness of the likely effects of these structural changes on the subject-access system is critical.

Considerable work has been done, mostly several years ago, on the evaluation of subject-access systems from the standpoint of costs and/or effectiveness. New techniques and new systems will require new evaluations, utilizing different measures of effectiveness than those used for earlier systems. This issue is not strictly a research issue; it also includes cost-effectiveness as a criterion for evaluating subject-access systems.

Throughout the workshop it was noted that rapid changes are occurring. Many of these changes are precipitated by an outside element, that is, new technological developments. As in other areas, the potential impact of technology is considerable and needs to be the subject of careful study. Of particular interest are new subject-index systems which can be designed to take advantage of new technologies and the related effect of such systems on policies, users, etc. The impact of technology to date on existing subject-access systems should also be considered.

In some areas, changes may not be occurring as rapidly as is desirable due to inertia and/or the difficulty of making changes in complex systems. Yet, as already noted, new developments with the potential for improved user access are available, and will continue to become so. One need expressed by the workshop participants in this regard is the creation of an environment for initiating change.

Many of the issues cited by workshop participants call attention to specific areas of research needs, a number of which deal with overall system considerations on a national scale. Others mentioned the need for research and development to exploit new techniques and technologies. In contrast with these expressed needs, a lack of R&D support for subject-access methods and systems was noted as an issue.

A final group of issues related to the subject-access environment concerns the material available for coverage and its characteristics. The range of materials requiring subject access is considerable, and characteristics vary considerably. In addition, the degree of coverage of different groups of materials varies substantially. Singled out in this area by the workshop attendees were the humanities literature with its distinctive characteristics and numeric data collections which lack subject-access systems.

Issues Relating to Interfaces

System - system:

- Need for bridging techniques for subject-access schemes
- Need for descriptive register of data bases
- Relationships between library catalogs and online access services
- Need for comparative studies of subject-access alternatives

User - system:

- Diversity of user types, diversity of questions
- Aspects of questions:
  - models of query formulation
  - query expression
  - query characterization
  - query profile transformation
  - vocabulary browsing
  - nonsubject indicators
- Lack of application of user/use for system improvement
- Immediacy of access
- Roles of intermediaries

Two major kinds of subject-access interfaces were addressed by workshop participants: the complex interrelationships between different subject-access systems and the relationship of the systems to their users. Issues were raised in both areas.

Considering system interfaces, workshop participants noted several difficulties created on by the quantity and variety of available subject-access systems. Two issues dealt with the comparison of different systems -- specifically, the relationship between library catalogs and online access services, and subject-access alternatives in the context of the same user questions. The former issue raises several points of comparison, including that of manual versus online systems and book versus journal article coverage. While the performance of individual subject-access systems is frequently evaluated, a need for a comparative study of different systems was expressed.

Viewing the future as an environment in which many subject-access systems will be directly available to the user in an online mode raises issues of how users will select among the available data bases and how their searches will take into account the widely varying languages and search systems which make up today's subject-access system. To help in the selection of data bases, a descriptive register of data bases, possibly online, was indicated as desirable. Uniformity in the language area was generally rejected in favor of bridging techniques which would allow a kind of translation among subject-access schemes.

In the area of user-system interfaces, thought was given to the range of users, including both the end user and intermediaries. Classes of users were mentioned -- the general population, scientists, researchers in the humanities, educators, business people, special interest groups, and so on -- and an even greater diversity of questions was seen as a fact of life to be responded to. A number of issues were raised concerning the nature of these questions, including the need for models of query formulation, query expression, query characterization, and query-to-profile transformation. Related issues will require the exploration of search techniques such as vocabulary browsing and the use of nonsubject indicators. Immediacy of access with online systems was noted, and the effect of this on the user was deemed to be an issue. Finally, development work leading to full use of user feedback in system design was called for.

After the identification of subject-access issues, the workshop participants split into small groups to develop suggestions for recommendations. Each group brought back a set of draft recommendations to the full workshop, and these were briefly discussed. Many recommendations, some overlapping, were presented and it was left to the Planning Committee to build upon the sense of the meeting to develop the final summary recommendations.

DRAFT RECOMMENDATIONS OF DISCUSSION GROUPS

Each of the five groups produced a set of recommendations. These were brought to the total workshop for discussion. Below are the statements of from all groups.

Group A

1. Investigate the design characteristics of a dynamic file system based on file-usage data from the user population and from collection change and growth. The system itself should be time independent, adapting to changing conditions through accumulation and organization of information to guide its functioning. Evaluation will include estimates of performance and cost of the file systems.

2. Examine the comprehensive and macro-level relationships of the various subject-access approaches to the information sought. A pilot project should investigate the response and performance of various systems by taking a few selected questions and following through the card catalog subject headings, a vertical file, abstracting and indexing services to journal articles, back-of-the-book indexes, etc. The search trail would be traced; subject entry points would be compared; the amount of material scanned would be recorded; and the results of each search would be compared. Two questions to be answered are "What information does the user take to the shelf in a library to identify material sought?" and "How many items have to be scanned before the information sought is found?"

3. Analyze the question-asking process from concept formulation to the interrogation of the system, including the formulation and expression of questions and the reformulation of selected questions in light of the system's response.

4. Construct a subject-access system incorporating comprehensive nonsubject indicators (e.g., journal title, publisher, author, etc.) for monographs and other literature similar to the combined searching approaches available through automated abstracting and indexing services. Research has indicated that searchers rely on such data as publisher or journal title as quality indicators.

Group B

1. Devote effort to exploring the technical and economic feasibility and utility of other approaches to multifile subject access rather than support continuing development of broad-based universal subject description schemes and automatic subject-switching mechanisms. Among other approaches include:

   a. the building of composite indexing records from multiple sources,
   b. developing publicly available retrieval aids (e.g., stored profile parameters).

2. Encourage increased cooperative efforts between and among A&I services and the national libraries to harmonize subject-access vocabulary in areas of overlapping interest.

3. Investigate the complementarity of subject access to monographs as currently applied by the national libraries and the A&I services, via an applied research project.

4. Make increased use of user feedback and data available from retrieval services to improve the design of data bases, through applied research projects.

Group C

1. Mobilize funding support to assist the Library of Congress to:

   a. create an online system to manage and distribute subject authorities,
   b. undertake necessary vocabulary reform in the list of LC Subject Headings and alter practices in assigning subject headings to works cataloged,
   c. eliminate subject cataloging backlogs so that attention can be turned to making subject changes without the pressure of undone work hampering progress in this area.

2. Fund the creation of a descriptive register of data bases, with a statement of the scope of each.

3. Seek funding so that the Library of Congress can provide increased subject access to audiovisual materials by:

   a. adding cataloged audiovisual entries to the LC MARC data base,
   b. extending the Cataloging-in-Publication Program to audiovisual materials,
   c. expanding the number of subject (and other) access points to audiovisual materials.

4. Fund a comprehensive study of various levels of use including:

   a. the behavior of users of card catalogs and other methods of subject access at different types of institutions (public libraries, research libraries, etc.);

b. barriers to subject access experienced by various categories of users, e.g., researchers, users of archives, public libraries, etc., and categories of users such as the disadvantaged, imprisoned, etc.

5. Investigate the provision of multiple levels of access to the same materials, providing funding for such a study and then funding for implementation if this is determined to be feasible. For instance:

a. investigate the feasibility of providing of nonresearch subject headings to public libraries;
b. provide cross linkages between vocabularies, classification schemes, and other taxonomical access tools, such as linking terms in Sears, LCSH, and ERIC.
c. support and accommodate special refinements of subject-access tools such as special classifications.

Two additional suggestions that were discussed in the group but not included in the recommendations above were:

1. Produce an illustrated manual on methods of making subject heading changes in card catalogs.
2. Hold a conference/workshop to study the possibility of creating a subject heading code (rules).

Policy statements agreed on by the group:

1. Standardization of vocabularies, classification schemes and access methods is a non-goal.
2. All types of publicly accessible information services must be included in considerations of subject access: public libraries, academic libraries, research libraries, archives, special libraries, etc.

Group D

1. With the desire to encourage the diversity that will best serve the end user and in response to the problems perceived in subject access, it was recommended that subject access should be the responsibility of each discipline or type of library. New organizational structures as needed should be developed through already existing A&I services or through other professional specialist groups. Zero-base thinking should be used to spur innovation. In other words, start from scratch so that advanced techniques such as post-controlled vocabulary and automatic indexing can be rapidly utilized. This approach can be of particular value to the humanities.

2. In order to encourage an environment for change it is suggested that the library community act as user intermediary to the agencies that produce information, e.g., the A&I services, index publishers, and other publishers by focusing on end user needs and informing those agencies of those needs. The expense involved in change may then be justified. With only an occasional complaint or suggestion from the marketplace the producers cannot evaluate change to benefit the user.

Group E

1. Perform research on the several general problems with subject access in the humanities that may be different from those in the sciences and engineering. One problem is that not much information is available on how humanists do subject searching. There is some question as to whether some notions about the potential value of A&I services are based on fact or on speculation and hearsay.

2. Retain diversity in subject-access mechanisms. We are unable to anticipate all the needs of individuals seeking subject-related information in all disciplines.

3. Encourage libraries that require subject access in greater depth than that provided by LC to band together and pool their resources.

4. Make direct access by users, insofar as possible, the goal of subject-access systems. This should not be restricted to online systems but should also include manual systems such as card catalogs; that is, subject-access systems should be made as simple and user interactive as possible.

5. Establish basic standards for the bibliographic identification of documents.

The recommendations that emanated from the groups were discussed by the full meeting. A summary of the recommendations is provided below for completeness:

1. Support and endorse a diversity of subject-access approaches.

2. Make simplicity in the use of a subject-access system a general goal, but the intended audience should be a guide for inclusion of specific techniques that will be most effective.

3. Make access available both through controlled vocabularies and natural language, where possible.

4. Establish a constantly updated directory of available data bases including access, scope, and cost.

5. Encourage library schools to include courses on subject access beyond the technical facility of subject headings, and extend such programs as part of continuing education.

6. Perform research to determine the basis on which records might be transferred from online status to some type of record storage, considering such factors as age, frequency of use, and subject field for the purpose of low priority retrieval.

7. Encourage NEH and other appropriate agencies to support research aimed at identifying the differing subject-access needs of the humanities as compared to other disciplines with the aim of promoting the development of new subject-access tools in areas where there are now none and improving those that are unsatisfactory.

SYNTHESIS AND ACTION RECOMMENDATIONS

Introduction

Building on the experiences of planning the workshop, the discussions that took place, and the recommendations suggested, the planning committee has constructed a limited number of action recommendations as presented below. To the extent that the participants contributed directly to these results through the workshop and follow-up correspondence, full credit for their insight and judgment is intended. However, to the extent that the planning committee has selected and combined ideas, placed emphasis, and otherwise shaped the action recommendations below, it has done so on its own responsibility and accepts any criticism that may accrue. It does not wish to imply a complete consensus with this synthesis by all the participants in the workshop, but the planning committee does hope that it has correctly understood the results of the workshop and expressed views and actions that most of the participants will support by their own actions.

In addition, recommendations recorded at the workshop but not subsumed in the planning committee synthesis contain other actionable ideas (see preceding section of the report on draft recommendations). The planning committee invites those who have a specific interest in aspects that were not selected for emphasis to take the initiative in developing them.

Major Themes

A few themes relating to subject-access techniques and activities stood out strongly during the workshop. The most overwhelming of these was "diversity," even though one of the intentions of the workshop was to seek "useful commonality" of problems, approaches, and resource sharing. Moreover, there are

serious communication and understanding gaps between the practitioners in the various communities of interest.

It is clear that diversity in subject-access needs and approaches

. is real,
. is good,
. will continue.

Diversity comes from the many environments of use of information and the varying structure and culture of subject areas or disciplines. Work to improve subject access must take into account the fact that there is no one way that can be applied to all areas. However, there are useful linkages among approaches and activities that appear desirable but which do not yet exist. Helping to highlight these linkages was one intention and theme of the workshop.

To illustrate sources of diversity: groups of institutions serving the general public have needs for subject-access mechanisms much different from those of major research libraries. Those needs are determined by the purposes to be served; the age, educational level, and culture of the patrons; the location of the population being served and the logistics of serving it; and other special requirements.

Another strong theme was that subject-access mechanisms should be designed for end users and not exclusively for information professionals -- whether the subject-access tool is a card catalog or an interactive computer search system. This is not to say that professional intermediaries should not or will not be employed in many endeavors. Users should not and need not be intimidated by the language and complexities of formal information tools nor be inhibited by the limited population of information professionals.

Consistent with the emphasis on adaptation to diverse needs and environments, "postcoordination" was emphasized over "precoordination" in the design of information-access tools. Those who organize information for retrieval can not anticipate

all future interests in and uses of the information. Thus, at least some component of the overall storage and retrieval system (and some would say, the major component) must be a way to determine at the time of retrieval relevancies not reflected in the original organization of the collection being searched. The possibilities for achieving this are improving with modern computer-based systems.

Groups having common interests must band together for mutual benefit, and resource sharing or other unified actions must go well beyond the boundaries of libraries to include publishers, abstracting and indexing organizations, and other information agencies.

An area that regards itself as deficient in organized subject-access tools, such as the humanities, can learn from other areas, such as science and technology, that have put more effort into development of such tools. However, the techniques developed for the sciences should not be accepted blindly; all may not be relevant to the unique information-seeking needs and conventions of the humanities.

Yet another theme was that universal standardization should not be an end in itself. Carefully selected standardization can be a useful, even necessary, tool in the reduction of redundant effort; and local standardization (by subject area or use environment) is a tool in effective system design. However, interlinkage among the various communities of interest (e.g. among the various vocabularies of subject cataloging and indexing) is far more practical and acceptable, and therefore possibly achievable, than any attempt at absolute uniformity.

Finally, a single grand plan with centralized funding cannot be expected. Whether or not a logical case can be made for a single national plan, it simply will not happen in the pluralistic society and economy of the United States. However, techniques and technology developed in or applied to one area may be of benefit to other areas, and furthering the transfer of useful technology is in the national interest.

Action Recommendations

The following action recommendations each require an agent for an organizing initiative, a funding source (possibly separate), and a set of people and organizations to do the work. The planning committee does not assume the authority to designate such participants by listing them in this report. Real progress can be accomplished only if those who have the capability to perform the work, those who can benefit, and those who can facilitate progress through providing study and development funds will step forward and take initiative and responsibility. By stating desired and specific actions in the paragraphs that follow, the planning committee hopes to provide a catalyst for that to happen.

1. Interlinkage of library and A&I produced subject-access tools

   Problem: How to bridge between library catalogs and subject-access tools and approaches produced by abstracting and indexing organizations (or other specialized information access organizations) to include:
   - . both printed and computerized mechanisms,
   - . both present techniques and evolution toward future possibilities,
   - . educational efforts in library and information science schools and in-service training,
   - . work tools needed to apply the techniques,
   - . integration of mechanisms both by the creator of subject-access tools and at the point of use of disparate tools.

   Actions: a. Locate action agents in each of the library and A&I communities who will collaborate to develop demonstration projects in a specific limited subject area in both of the categories of

"hard" science (e.g. physics, chemistry, biology, mathematics, engineering, etc.) and "soft" science (e.g. psychology, sociology, anthropology, etc.). The results of the demonstration projects should be an understanding of the specific subject-access problems of the selected area, a solution of them in an integrated environment, and pilot experience to guide subsequent development in other subject areas. The specific focus is on how to bring library and A&I tools together, not on the development or exploitation of either alone.

b. In a similar manner, plan and carry out a demonstration project on the linkage between formal catalogs and access tools to manuscripts and other forms of information in a specific area of interest to the humanities.

c. At the beginning of each project, require the assembly and dissemination of a state-of-the-art review as it relates to the subject access in the chosen limited subject area, as well as require full documentation and dissemination of the final results of the project.

d. Additional integrative efforts are needed to broaden awareness and understanding of the various kinds of subject-access techniques and tools. At present, publication and professional discussion of these subjects tend to be parochial and limited. Toward this end:

. establish joint sessions and/or special conferences sponsored by professional associations (e.g. ALA, ASIS, SLA, American Society of Indexers, etc.) on current techniques and developments in subject access. Structure such conferences in such a

way as to break down communication barriers among practitioners in the various professions (e.g. subject catalogers, indexers, reference librarians, information specialists, information system researchers, etc.)

- Prepare publications, including up-to-date textbooks and curriculum materials, treating subject access in an integrative sense for formal education and training purposes.

2. Toward a System view of future subject-access mechanisms

Problem: How to design a large subject-access system that adapts to usage patterns and collection change and growth without requiring disruptive alterations in system structure or costly file conversions. Motivations for solving this problem come from the factors that are causing the closing of large card catalogs, the difficulties of weeding out inactive material from large collections, and the changes in vocabulary and concepts of interest as time passes.

Action: Conduct a research project on the design characteristics of a large dynamic file system based on file usage data from the user population and from collection change and growth. Collections in excess of one million active items requiring files in excess of one billion bytes are considered large for this purpose. The project should include:

- the study of use patterns and user behavior in subject access at different types of institutions,
- collection-purging criteria and phenomena,
- consideration of a user- (or use) created

cross-reference system,

- theoretical characterization and analysis,
- synthesis into a system design with analysis of its dynamics and economics,
- application to one or more demonstration projects in limited areas.

The results should have implications for:

- design of future large systems,
- decisions on selective conversion of back files to new systems,
- establishment of subject access for audio-visual and machine-readable files not now satisfactorily cataloged and indexed for subject retrieval,
- adaptation to end user needs.

3. Gaps in subject-access tools for the general population

Problem: How to bridge gaps in the availability of subject-access mechanisms to serve special needs in school and public libraries. Great diversity of need arises from the broad range of age levels, socio-economic environment, language, and the special needs of the handicapped. Yet the extent and nature of these needs are not well understood by much of the information profession.

Action: Make and disseminate an analysis (description and evaluation) of the availability and suitability (usefulness/effectiveness for the audience) of subject-access approaches and tools for users of school and public libraries for all purposes (e.g. to include at least educational, personal, and community activities purposes). The study should include determining the "how" and the "why" of uses by each population segment.

4. Subject access for the humanities

Problem: How to respond to the general perception of those directly involved that subject-access tools for the humanities are incomplete and lacking in availability. Although certain quality tools do exist (e.g. Philosopher's Index, Arts-&-Humanities Citation Index), there are many unfulfilled needs, and methods used in other disciplines of knowledge are not as extensively applied in the Humanities.

Actions: a. Conduct research to determine why this situation exists that will study these questions:

- What is the nature of the humanities literature?
- How do humanists use information?
- What design of information tools is appropriate to respond to the characteristics of the literature of the humanities?

b. Set up orientation courses and follow-up task groups for humanists to review information techniques used in other fields, to critically evaluate them, and to explore how to apply the techniques, that are judged to be relevant to their needs.

5. Subject access to monographs

Problem: How to deal with the assertion that detailed subject access to monographs is inadequate via subject headings in card (or equivalent) catalogs and that other tools for that purpose generally do not exist. Recent studies show that more detailed subject access to monographs is potentially useful, at least for non-fiction works. Ideally, the analysis and derivation of

in-depth subject cataloging and indexing material should be done at the source (by author or publisher), but there is no economic incentive to do so. A separate analysis/indexing process in the manner of treatment of journal articles by abstracting and indexing organizations is uneconomic due to the time required for a subject expert to read the monograph and perform the analysis. The tables of contents and back-of-the-book indexes in many monographs are not exploited for this purpose. In addition to general information retrieval purposes, good subject access to monographs would enhance the effectiveness of blanket order plans with publishers by more specific screening to fit the needs of specific libraries. Many technical ideas that have been suggested at various times have failed for economic reasons.

Actions: a. The workshop generated no suggestions that would have widespread applicability for making significant headway on this problem. The situation remains an important unsolved problem to be faced by publishers, abstracting and indexing activities, and libraries, singly or together.

b. Without prejudging the outcome and without suggesting that the method would solve the entire problem, the following research project might lead to useful results for some future purposes. Study the generation of useful index material by statistical and/or linguistic analysis of the text of monographs available in machine-readable form as a by-product of computer-controlled composition. The conditions that lead to this recommendation are:

- . the increasing use computers by publishers,
- . the capacity and speed of computers are rising and costs are falling dramatically,
- . something useful should be better than nothing, as is the situation now; traditional subject headings or index entries need not be required.

6. Subject-access authority files

Problem: How to cope with the creation of a national library network which requires that subject-access mechanisms be established to work in that environment and to provide multi-institutional access. One key to the effective functions of such a system is an appropriate subject authority system of broad applicability.

Action: a. Include in the work toward an authority system for the Bibliographic Service Development Program means to link subject-access vocabulary authorities of both the library and the A&I communities and use the participation of both communities in developing that linkage.

b. As a related effort, assemble a work group to study the possibility of defining codes (as in "cataloging code" or "indexing rules") to guide the construction and use of each subject-access vocabulary included in the authority system. Work guides used in various subject areas by the corresponding cataloging or indexing agencies should provide much guidance for this work.

APPENDIX A

SUBJECT ACCESS IN PERSPECTIVE

Professor F. Wilfrid Lancaster

October 18, 1978

Thank you very much. I noticed when I began to prepare for this talk that its acronym is SAP. I am not sure whether that is significant or not. But in my preparation, I began by considering the problems involved in the design of effective subject access systems, and while I was thinking about the problems I was reminded of some work that I did approximately ten years ago when I was working for the National Library of Medicine and engaged in an evaluation of the MEDLARS service of the library. This evaluation involved the analysis of the performance of a rather large number of searches, approximately 300, that the National Library of Medicine undertook for their users. As you can imagine, some of these searches retrieved some completely irrelevant citations. Two of these searches became my all-time favorites and I do not believe I will ever forget them. I would like to begin this talk by sharing these with you.

One of the searches conducted was on the subject of separation anxiety in childhood, the objective being to retrieve information relating to the anxiety of a child at being separated from its parents. One of the completely irrelevant articles that was retrieved by this search had the delightful title of "Psychological Aspects of Circumcision". A second search in the same group was on the subject of premature rupture of the fetal membranes. The search on premature rupture of the fetal membranes retrieved a completely irrelevant document entitled "Project Headstart". Now I mention these two facts for two reasons. First of all I could not think of a better way of beginning the talk, and secondly, because these examples are indicative of the fact that even with highly trained indexers, a

carefully controlled vocabulary, and experienced searchers, information retrieval systems do not always retrieve what we want or expect them to retrieve, although for certain purposes they may retrieve something better.

I think everybody in this room would be willing to agree that we have seen very great improvements in methods of subject access since World War II. In this period there have been, I believe, four major developments. The first was the development of the first so-called coordinate or postcoordinate indexing systems, which allowed us to achieve for the first time freedom from the linear sequencing of subject terms that was characteristic of card catalogs and printed indexes; and indeed to achieve for the first time a true multi-dimensional approach to subject matter efficiently and economically. Later, we saw the progressively more sophisticated implementation of these postcoordinate systems through the principles of optical coincidence, edge-notched cards and microfilm; later, punched cards; and somewhat later still, through magnetic tape and magnetic disc. Thirdly, we've seen a quite spectacular growth in the number of machine-readable databases and data banks available for the provision of information services. These have grown from one to several hundred in approximately a fifteen year period. Fourthly, we have seen an equally rapid growth of computing and telecommunications facilities for making these resources accessible widely and inexpensively online.

The period since the Rapid Selector and since the work of Batten in England and Calvin Mooers in the United States in the 1940's has been one of rapid change in methods of subject access. It has been a period in which wheels have been discovered, rediscovered, forgotten and discovered again. It's a period in which we have gone from controlled vocabularies, to natural language, to controlled vocabularies and back again to natural language. It's a period in which we've gone from systems, namely card catalogs and printed indexes, designed to be used by more or

less anyone, to more esoteric systems designed only to be used by information specialists. Now we are trying to rediscover how to design effective "anyone systems" again. What have we learned in this period? What subject access truths have emerged? What knowledge has been gained to guide information systems design in the future? What are the gaps in our knowledge? What are the neglected areas of research? Let me mention some of what I feel to be the major things that we've discovered in the last few years.

Subject access systems are complex organisms. A succession of evaluation studies has revealed that there are indeed many factors which determine the success or failure of a search in such a system. These factors include: the coverage of the database; the indexing policies of the database producer; the quality and consistency of indexing in that database; the effectiveness of the interaction between the users and the system; the completeness of the searching strategies; and the characteristics of the vocabulary used in the indexing and searching operations.

One of the problems, and in some sense the major problem, faced in the provision of information service, is that most of these factors are not directly under the control of most information centers. In fact, in a typical information service situation, the information center controls only a few of the factors governing its own performance. Consider, for example, an industrial library searching machine-readable databases on behalf of its engineer and scientist users through some online service center -- I'll use Lockheed as an example in deference to Charlie Bourne -- through Lockheed. The library itself controls selection of the database or databases it is going to search, but the staff of the local library has no control whatsoever over the quality of the database it uses. It has no control over its coverage; it has no control over its indexing policies; it has no control over the accuracy of the indexing; and it has no control

over the vocabulary used to represent the subject matter. It has some control over its own search strategies but not even complete control over these because the search strategies must be constrained by the characteristics of the database itself and by what the system software, the query language in this case, allows the searcher to do. The local information center has most control over the methods by which it interacts with its own users. But even here the control is not absolute since it is impossible to exercise complete control over users and their behavior. You can lead a horse to water but you can't make him (or her) drink.

Another thing that has emerged rather clearly in the last few years is that it is rather difficult, in fact extremely difficult, to evaluate the performance of subject access systems. The results of a known item search can be judged as either successful or not. But the results of a subject search cannot be evaluated on a simple binary scale. They must be assessed on some form of graduated scale which takes into account how much of what is wanted is retrieved and how much of what is retrieved is wanted. Such a scale implies the use of values, such as the values of relevance or pertinence, which tend to be subjective, transient and relative rather than objective, stable, and absolute.

Another thing that I believe we've discovered is that there is no single correct approach to the design and implementation of information retrieval systems. A system capable of performing at some specified level can be implemented in several possible ways. In this connection, it is important to recognize the obvious trade-off between input and output costs in the operation of information services. We can choose to deliberately reduce the input costs or effort, for example by adopting some minimal level of indexing such as keyword-in-context, but only at the expense of increased cost and effort at the time of the search for the information. This leads me to the related matter of

standardization. Efficient access to documents by author or title requires standardization. The fact that author/title cataloging can be reduced to standard rules equally applicable in all libraries means that these activities can be and should be handled in a centralized or cooperative mode. But we must be very careful when we talk about standardization applied to subject access.

In the notes prepared by the organizers of this meeting an important question is raised as follows: "How can we eliminate expensive and time-consuming duplication of intellectual processes in creating subject access to bibliographic entities?" My answer to this is that I do not believe we can, at least not by using any conventional approach to the problem. Subject indexing and subject cataloging are not susceptible to standardization and centralization. Indeed, standardization and centralization may be regarded as the natural enemies of effective subject access.

Effective subject indexing involves the establishing of answers to two questions. The first question, What is this document about? The second question, Why are we interested in it? Question number one may be handled in some standardized, centralized way. Question number two cannot by any stretch of the imagination be centralized or standardized. Ten different organizations may all agree on what a particular document is about, but each may have completely different interests in this document and completely different reasons for adding it to their collections. They have, therefore, completely different needs for subject access to this document. We should not be disturbed or even surprised to find that the same document has been indexed in ten different ways, in ten different information centers. Indeed the reverse is true. It would be cause for surprise and for concern if we did find the identical indexing of this document in all ten organizations.

Another point, we can recognize over the years a move to

simplification in the design and implementation of information retrieval systems. The systems of today, although implemented on more sophisticated equipment, are in many ways simpler than many in use in the 1960's. Many of the earlier systems were over-designed, incorporating, for example, links, roles, relational indicators, semantic codes and other devices to filter out noise or to achieve more complete search results. These refinements could rarely be justified from a cost-effectiveness viewpoint. They gave marginal improvements in performance and these marginal improvements were not enough to offset the greatly increased costs at input.

Information retrieval now seems to be looked at much more pragmatically than it was twenty years ago. There seems now to be more realism in our approach to information retrieval problems. Writers twenty years ago were very much concerned with identifying possible semantic or syntactic problems which might affect the retrieval process. Most of the problems that they unearthed were more theoretical than practical. For example, is it likely that anyone will be looking for information on blind Venetians? If so, is it likely that information on blind Venetians will appear in the same database as information on venetian blinds. And if it does, how much effort is needed to segregate the blind Venetians from the venetian blinds after the search is completed?

It is sometimes overlooked that words that are quite ambiguous in isolation may be quite unambiguous within the context of other words. Information retrieval is rarely concerned with words in isolation.

Finally, I think, in terms of our discoveries over the last few years, it has been recognized that it may be possible to provide effective subject access through approaches that are somewhat unconventional, including approaches that may be partially or completely independent of the vagaries of language. These approaches include the approaches of citation indexing,

bibliographic coupling, and some form of pattern matching whereby we can ask a retrieval system to find other documents that are in some sense like a document we already know to be relevant to our interests.

We have learned a lot about subject access since the 1940s. While we have not found all the answers, at least we may have identified some of the more important questions. However, much still remains to be learned. There are many gaps in our knowledge and many research areas have been neglected. For example, the most critical aspect of an information service is that of the interface between the information service and its users. When I talk about the user system interface I do not necessarily mean man/machine interaction. In fact, I am more likely to mean person to person interaction: the interface between a user with some information need and a member of some information staff who will attempt to satisfy this need. I refer to this as a critical area since it is here that a latent need is transformed into an expressed need, that is, a request for service from the information center. Clearly, the whole information retrieval operation depends upon the success of this transformation, that is, the extent to which the expressed need accurately reflects the actual need. We may have a system in which documents are indexed exhaustively, the quality of the indexing is carefully controlled, the vocabulary is carefully constructed and sufficiently specific, and the search strategies are comprehensive; but all of this matters little if in fact we are looking for information our users neither need nor want.

The user/system interface is the most critical component of an information retrieval system. It is also in many ways the most neglected. We still do not know for sure what methods of interaction are most effective in capturing requests that are accurate representations of the needs that underlie them. It is at the user/system interface that most operating information services can achieve the most significant improvements in their

own performance.

We do not know a great deal either about the language with which users approach subject catalogs and subject indexes. From where do they draw their search terms? What are the factors that influence the user's choice of entry terms? Symptomatic of neglect in this area is the fact that most computer-based systems have traditionally maintained records of the frequency with which terms are used in indexing; but how many have routinely maintained records of the frequency with which terms are used in searching?

In the 1960's considerable experience was accumulated in the design of computerbased retrieval systems. But most of this experience was experience related to the design of systems to be used by information specialists. Now we are faced with the design of systems to be used in a non-delegated mode by chemists, physicists, economists, and other professional practitioners. This is a completely different kettle of fish. Should we design such systems so that the workings of these systems, rather like the automatic transmission in an automobile, are completely transparent to the user, or should we design them with sensitive and high performance gear boxes which are entirely under user control? Is it reasonable to expect the infrequent system user to learn the nuances or, I would even say, idiosyncracies of a controlled vocabulary? Or is it more reasonable to design the system in such a way that it can be interrogated in the user's own uncontrolled vocabulary--the natural language of his subject field?

What of the future of subject access? The changes of the next forty years will surely be more dramatic than those of the past forty. The entire structure of the provision of information services, especially the economics of these services, is changing. Indeed the entire structure of formal and informal communication may be changing. We no longer need to think of a major reference library as a large collection of print enclosed

within four wall. We no longer need to make a capital investment in a database to make it physically accessible, that is, we no longer need to have it sitting on the shelves of a library. Information services are moving rapidly into an on-demand, pay-as-you-go environment; an environment in which one can purchase access to whatever information sources are needed at the time they are needed. This situation greatly improves the accessibility of information sources, since a database needed once a year is as accessible physically, if not intellectually, as one needed a hundred times a year.

So far most of the use of online services has been use in a delegated or mediated mode. This will undoubtedly change as terminals fall in price and, as a result, become increasingly available. The problems of subject access in the future then are likely to be related to the design and implementation of user oriented systems. Systems whereby the practitioner in some field, who is not an information specialist, may use a terminal to access an almost unlimited array of information sources. What are the problems of operation in this type of environment?

The first problem may simply be that of an embarras de richesses. So much will be accessible through a single terminal that the user will be overwhelmed with the resources available. The first requirement will be for some form of automatic referral mechanism whereby a statement of the user's need in natural language form can be matched against some online directory of available resources. The result of the match will be an identification of those sources which have the highest probability of satisfying this need together with an indication of how these sources can be accessed and perhaps how much it costs to access them. In this environment in which many different databases may be accessible from a single terminal, much use of databases will be made by the infrequent rather than the regular user. This may mean a continuing move toward natural language and away from controlled vocabularies, at least

controlled vocabularies in the conventional sense, since natural language systems are generally more hospitable to the infrequent user and since a query in natural language form can readily be transferred from one natural language database to another.

Does this then imply a declining interest in thesauri and vocabulary construction in general? Not necessarily so. A controlled vocabulary, as you know, has two major purposes. It exists to normalize the terminology of a subject field and it exists to provide a structure whereby terms whose meanings are related are brought together in some way. It has generally been assumed that these purposes must be served by a precontrolled vocabulary, a vocabulary in which this normalization and structure is imposed at the time of indexing. But we are not necessarily restricted to precontrolled vocabularies. There is no reason why we cannot develop postcontrolled vocabularies--vocabularies that are tools only of the searcher of natural language databases; tools such as synonym tables and/or loosely structured thesaurus groups that are constructed by humans, by computers, or by a combination of man and machine. The future, I believe, lies in this direction. Through the use of natural language input, and some type of postcontrolled vocabulary at output we can combine at least some of the advantages of both into a single system.

In conclusion, in your deliberations during the next two days, I would urge that you look at subject access in terms not of the past and not even so much of the present, but in terms of the needs of subject access in a future electronic, on-demand, pay-as-you-go society. Thank you very much.

APPENDIX B

SUBJECT ACCESS: CHALLENGES FOR THE FUTURE

Margaret T. Fischer

October 19, 1978

In college a very special professor was trying to instill the importance of good organization, dynamics, and quick thinking. It was at the end of a writing course when he said to us, "I'm going to give you four things to write about in forty minutes. Tell me about mystery, royalty, life and religion." Not many people did much for the first few minutes. Suddenly a student got up, went forward and handed a piece of paper to the professor. He started to walk out but the professor said, "Wait, you've got to try." The student said, "I did. You told us you wanted organization, you wanted us to do it quickly, and you wanted dynamics." The professor picked up the paper and read: "My God, said the princess, I think I'm pregnant. I wonder who done it." Here are the words -- not in isolation as Wilfrid Lancaster said last night. They are curt, they are clear, and they are concise. I would hate to index this little scenario and there would be delightfully unexpected smiles at retrieval. But it illustrates the heart of our problem -- words.

This simple example stresses the information process as being part of the process of discovery and of innovation. It asserts that the essence of the information problem is to maintain knowledge as a viable unity and that the basic information processes are those of selecting, of reviewing, and of synthesizing information. Perhaps we should think no more seriously about a fully automated information system than we would think of a time and discovery machine, but instead to work toward bringing collective wisdom to bear on the everyday management of information using as much relevant technology as possible, yet mindful that our system will be built in levels,

some more beautifully conceived than others, rather like a latter day hanging garden of Babylon.

Where do we begin with the Alpha and the Omega? What is this all about? Why are we interested in it? Since every collections manager will answer differently we said last night, let the publisher in concert with the author begin to give us the answers of the Alpha "in which our hero does so and so," said David Batty; "patchwork quilt," said Lucille Gordon; "all for better information;" "if that would speed delivery I'll do it;" and "I'd rather work with book in hand." Being a publisher I'd rather see us begin with publishers. What better place to know what this information is, what it is all about, and why we are interested in it! Publication has been extended far beyond our present ability to make real use of the record, so let the record through the publisher begin to make use of itself.

Our first challenge, it would seem, is to work out a way in harmony with publishers, not only publishers of the original materials but publishers of the surrogate materials. They are sellers -- purveyors of information. We have heard some of them speak during this last day and a half. They must make money in order to survive. They are as interested as we in providing the keys that will allow others to understand the value of the information which they purvey. I firmly believe that we can work with publishers as they develop the means of selling their information and at the same time begin to index it. The Alpha. We've got to go on to say, as my parents taught me for years, God give me the grace to accept what I cannot change and to change what I cannot accept. What we must accept are some of the truisms that we have been speaking to each other about. "Inconsistency," we say as if it were a bad word. "Inconsistency in indexing." So be it! I will accept the Tefko Saracevices and the other researchers who tell us by their wisdom and the dint of their effort that indexing is inconsistent. Perhaps we should say consistency is the hobgoblin of big minds. And so let us

accept inconsistency. The user -- we are so concerned about the user -- the end user, the immediate user, the intermediate user. The user is the user is the user! Whoever seeks information, wherever that place may be, is the user. Why not accept what we cannot change and get on with changing.

There are times when publishers index or abstract for us, yet none of us ever use the work. For years, the New York Times has had the essence of the news abstracted on the first page of the second section. The Wall Street Journal prints the key stories in capsule form magnificiently displayed in the second column on the lefthand side for all who wish to know about business information. I don't know anyone who consciously makes use of these abstracts, including people who work at the New York Times Data Bank and the Dow-Jones Information System. Let us open our eyes and look and see what is there.

Pictures -- I worked for twelve years for LIFE Magazine -- pictures are worth a thousand words. Do we use pictures? Eugene Garfield learned how to use pictures without ever taking one. What Gene saw in his mind's eye were pictures of contents pages, although he "photographed" them in a different form. But come back to thinking about snapshots. Bottom-of-the-line television equipment, the very cheapest set with the worst resolution, can transmit sixteen reasonably good pictures a second. The record is made by a moving beam of electrons sweeping across a picture very rapidly and projected upon a screen which glows momentarily when the electrons hit. The human eye can absorb that and more. Think if, for example, we took a picture of the cover of the contents page of a monograph, and of what we were going to index. It is not very costly because television is very inexpensive. Sixteen frames every second; we can look at them almost subliminally or we can look at them more slowly because we already have the technology to monitor at a slower scan rate. What I am getting at is that the picture is finished as soon as it is taken. The Polaroid people have been telling us this for

more than twenty years. Let us, as we ponder, think of the myriad of ways in which the picture is finished as soon as it is taken and think of that picture as information. This means we should try to have access assured at the point of publication in any form; and there are many, many forms.

The Omega -- All afternoon I listened; not once did anyone say, "information." Why are we in business? To provide information on all levels, in all decks, to all kinds of people. That means that we must try to understand the roots by which information comes to us. I suspect we know somehow that facts are easier to deal with than ideas, that entertainment as in the Gothic novel is easier than knowledge, and that static, historical information is easier than volatile, everchanging information. But information -- not indexing, not cataloging, not faceting -- information and the means of getting at it is our business. I contend that people know how to get information when they need it badly enough regardless of what the system is. If the system is really bad, they will not use it. They will find their own way even if it is through the "College of Colleagues." We will get what we need when we need it some way. Perhaps above all we should remember that.

If we are going to get information, the most important thing to seek is answers. Certainly we all know that answers have to do with postcontrolled vocabulary, with the vocabulary of output, and with online answering. In the business of analyzing answers, we must of course know the grand question, the save search strategy, the number of words used in searching, the natural language input, and the library of search strategies -- all these things we have been talking to each other about for this day and a half. It is the question that produces the answer and the answer that manufactures the question. They are one and the same.

I once had an experience with the business of questions and answers. We collected and analyzed 15,000 questions, most of

which came in via the telephone. We learned two essential things because we were able to put these questions to machine-readable files of abstracts, text, reports, and a vast array of junk, as Ed Papenfuse would call it. Sometimes like the girl with the curl we did very, very well, and sometimes we did horribly. We learned first that by and large the question is negotiated because the asker does not know what he wants. The question is usually the beginning of a question as for example, "I want to know all about China." The user does not want to know all about China because we would have to back up a truck and fill it. "Well, now, I do not really want to know all about China, I want to know about" -- and the dialogue begins. It takes a long time before the dialogue ends. Thus we talk about the problems of search strategy; the problems in classification, cataloging, and indexing. Number two is another of the givens in which we ask God to give us the grace to accept what we cannot change, namely, that there must be different schemes, schedules, classifications and levels for different needs. I may be very wrong but I cannot see how we can serve a world of users with one way. The way to develop the various schemes is to know the market. Not too many of us here concentrate on markets but publishers certainly do. You've got to know your potential user. If you want to expand your business you must know to whom you are expanding it. What are their questions? What are their answers? What are their needs? Who are they? Where are they? Knowing the market is knowing how to devise your product. It also tells you something about the kinds of answers you should provide. Some say they want real answers -- not pointers to answers. They want data -- not bibliographic citations. They want text -- not abstracts. They want abstracts -- not expanded titles. But the truth is that sometimes each of these things serves its purpose to get at the information. If what I want is a book on the renaissance, I really do want to use a card catalog. The catalog gives me the route to my answer. And so let us not throw out these various ways of getting at information.

For example, Dow-Jones has a new information service fashioned for people who only want three months worth of data because the users are interested in current information about business; about the stock market; about stocks and bonds; about people in business; and most of all about corporations and mergers and acquisitions. This information is so highly volatile that Dow is not going to go back retrospectively. They are carving out a special market for current information.

The people at the New York Times say, "They're crazy, the only way to go with news information is retrospectively." The New York Times is penetrating another market.

In the beginning, Dow employees that they did not need full text so they cut stories at the bottom and excluded some short stories. Dow soon discovered that its customers demanded every single word that has appeared in current Barrons and the Wall Street Journal. Users want every word because sometimes every word counts. Full text is required. Conversely, people at the New York Times say seventy to eighty percent of the answers that their customers need are included in the abstracts. Ergo there is no universal way to offer news to the marketplace.

Now we all know that a record, if it is to be useful, must be continuously extended. It must be stored in a way that it can be found and above all it must be consulted sometime or another. If we are going to extend records or change records, we must also "deep six" them or the seas of information will swell beyond the limits of good fishing. This is something else we have not really come to grips with -- how to "deep six" a record. We are fearful. Mendel's concept of the laws of genetics was lost to the world for a whole generation because his publication did not reach the few people who were capable of grasping it and extending it. We fear that this sort of catastrophe will be repeated; that truly significant attainments will become lost either in the mass of the inconsequential or in the mass of what we do not keep. And yet, if we are going to serve our market, we

must learn not only how to select but how to reject later.

Creative thought or volatile information, and repetitive thought or static, historical information are really very different things. For the latter, there are and may be powerful electronic aids. The repetitive processes of thought, like repetitive uses of information, can be regulated and defined like matters of arithmetic, statistics and tables. In fact, every time one combines and records facts in accordance with established logical processes, the creative aspect of thinking is concerned only with the selection of the data and the process to be employed -- meaning, "which one do I want?" But subject recognition and manipulation thereafter are repetitive in nature and hence a fit matter for us to think about relegating to the machine. This business brings us to the thinking that subject access should possibly present subject recognition in the sense that we had pattern recognition developed by machines. It begs us to look for machines and their programs for advanced analysis. True, the users of advanced manipulation and advanced methods of manipulating data are very small in number indeed. But some very exciting things are happening. There are machines for forecasting and report generation which take information about an industry or specific information about a corporation to play "what if" games, or to compare the company's performance against the national norm. There are machines for solving differential equations, functional and integral equations, and for that matter there are machines like the harmonic synthesizer which predict the tides.

Now I am not hung up on machines. What I am hung up on is how we can possibly think about using our own minds and the capability of the machine to do something that we have not been able to do before. Within the last ten years a whole new group of people has devised a kind of middle man of all get out. We call them the purveyors of numerical databases, a misnomer. Some of these databases are numerical in content, others are mainly

substantive and some are mixed, but all can be called information databases because all the information is there. You need go no further. What these purveyors have done with no bibliographic control whatsoever is to provide subject access to industries by recognizing two things: 1) they know how the people in that industry think and the words that they use in conducting their business so they use these words in searching; and 2) they know what data these corporations need in order to conduct their daily business life.

I am going to tell you about one purveyor that is only five years old -- not necessarily the most significant and certainly not the one that makes the most money at this point. This company is the epitome of what this new information business is about. Marine Management Systems is located in Stamford, Connecticut. Their system is called MARDATA. Five years ago the principal officers said to themselves, "We know the shipping business. Most of the information that shippers need sits in Lloyds of London, the Journal of Commerce and Shipping, charter brokers, the Tanker Advisory Center, and insurance companies that insure the ships that go about carrying the world's cargo. The world's cargo resides primarily in 30 corporations." Thirty corporations -- just the way 180 book publishers provide almost all the books that we Americans read -- thirty corporations provide virtually all the shipping of the world. So MMS set out to get these 30 corporations by providing them with all the computing facilities and databases which the maritime industry needs to move ships about, to man them, to give them food, to decide which ship is better equipped to carry cotton from Alexandria through the Canal and Red Sea to Sri Lanka or wherever. MMS developed the programs to do financial reporting, forecasting, planning, as well as operations. The programs include voyage analysis and reporting, economics and planning, vessel performance, transportation and cost analysis, cargo booking, terminal control, and general accounting. The users are the top executives in the shipping corporations. This is just

one example of a highly dynamic business being performed by people who never heard of cataloging, of indexing, of subject access, or of bibliographic control. Information for industry is being done now in advertising, in banking, in financial communities -- and we can and should learn lessons from this.

Relief must be secured from laborious, detailed cataloging, indexing and classifying even though surely we will be doing some of this for a long, long time. If the users are to be free to use their brains for something more than repetitive, detailed transformation, we must get on with our business. One might say that a mathematician, or a financier or a cataloger is not a person that can readily manipulate figures or words. How come? This is not even a person who can readily perform a transformation of equations for the use of calculus, or the transformation of words for the use of indexes. This person is primarily an individual who is skilled in the use of symbolic logic on a high plane. And especially, this is a person of intuitive judgment in the choice of manipulative processes that are employed. All else, this person should be able to turn over to some organization or some mechanism just as confidently as one turns over the propelling of a car to the intricate mechanism under the hood.

Whenever logical processes of thought are employed, that is, whenever thought for a time runs along an accepted groove, there is opportunity for better organization whether it be a catalog, an index, or a thesaurus. And there is opportunity to utilize technology to help drive it. Put a set of premises into such a device and an organization, turn the crank and it will pass from conclusion to conclusion. This is a much larger matter than merely the extraction of data for the purpose of research or seeking information. It involves the entire process by which man profits by his inheritance of acquired knowledge. The prime action of use is selection and here we are halting indeed. Our ineptitude in getting at the record is caused largely by the

artificiality of so many of the systems that we use for indexing, cataloging and fishing. We call it serendipity.

We place data in storage alphabetically or numerically and information is found by tracing it down from subclass to subclass. Information is generally in one place unless duplicates are used; and one has to have rules as to which path will locate it. The rules are cumbersome. We have talked all day about how cumbersome the rules are. Having found one item, save search or not, we have to emerge from the system and re-enter on a new path. The human mind does not work this way. It operates by association -- with one item in its grasp it snaps instantly to the next that is suggested by the association of thoughts in accordance with some intricate web of trails carried by the cells of the brain. Trails that are not frequently followed are prone to fade. Items are not fully permanent. The brain does get rid of stuff. Yet the speed of action, the intricacy of the trails and the detail of the mental pictures are awe inspiring beyond all else in nature. We try in our classification and in our thesaurus schemes but we have much more vertical association. Pretty much only the cross-references with which we grope haltingly, give us the horizontal association of trails.

Humans cannot fully hope to duplicate the mental process artificially, but they certainly ought to be able to learn from it. The first idea to be drawn from the analogy is selection -- selection by association rather than by indexing with clusters or facets. One cannot hope to equal the flexibility with which the mind follows an associative trail. But it should be possible to beat the mind decisively in regard to the permanence and the clarity of the items resurrected from storage. Little-used trails will not fade. What we need along with our association is a synthesizer. One would hope that we would use the reviews that we find in our literature, the committees for choice and selection, to begin the process of synthesizing knowledge. There

are two things that we need. There is no question, probably because of Mendel, that we need to keep what I call "raw data" somewhere -- the Library of Congress is as good a place as any. But the information with which most of us live in the sense of management systems -- giving data to a specific market for users to conduct their business -- should be synthesized in order for it to be used wisely and profitably. This means that we want things ready made with a mesh of associative trails running through them, ready to be dropped into storage and there amplified. The lawyer has his associated opinions and decisions; the patent attorney has millions of issued patents on call; the physician has the reactions of patients trailed to earlier and similar cases coupled with pertinent notes on anatomy and biology; the chemist has lists of organic compounds with trails to the analogies of compounds and side trails to their physical and chemical behavior; and the journalist has biographies of people and corporations with a large dower of history thrown in.

The challenge is to us -- we who are ever the professionals trail blazing in information -- people who find delight in the task of establishing useful trails through the enormous mass of the common record. I would like to think that the association aspect is like that old game we used to play as youngsters -- animal, vegetable or mineral -- very much like "What's My Line?" Remember how quickly and with what enormous facility -- using information theory, by cutting things in half -- we arrived with great speed at the answer. "Is it bigger than a breadbox?" All our steps, in creating or absorbing material of record, are perceived through one of our senses -- the tactile when we touch keys, the oral when we speak or listen, the visual when we read. Let us learn to use better the atificial senses of multi-media for subject access.

Humans have built a civilization so complex that we also need to organize records more fully if civilization is to advance to a logical conclusion and not bog down by overtaxing a man's

limited memory. Never has there been a statue erected to one who let well enough alone!

Now what is this path? It is the path of Aristotle. It is the path of Diderot. It is the path of Bacon. And most recently it is the path spelled out for us by H. G. Wells. I would like to read to you from a paper he wrote on November 20th, 1936, given at the Royal Institute of Great Britain at their weekly evening meeting. This is just a tiny, final portion of it. He speaks of the collective wisdom and knowledge of mankind and our need to synthesize it. Wells calls it a world encyclopedia. He does not mean the Britannica or anything you and I know. Listen to his words.

> "This world encyclopedia would be the mental background of every intelligent man in the world. It would be alive and growing and changing, continually under revision, extension, and replacement from the original thinkers in the world everywhere. Every university and research institution should be feeding it. Every fresh mind should be brought into contact with its standing editorial organization. And on the other hand, it's contents would be the standard source of material for the instructional side of school and college work for the verification of facts and the testing of statements everywhere in the world. Journalists would dain to use it. Even newspaper proprietors might be made to respect it. Such an encyclopedia would play the roll of an undogmatic bible to a world culture. It would do just what our scattered and disoriented intellectual organizations of today fall short of doing. It would hold the world together mentally; to hold men's minds together in something like a common interpretation of reality. There is no hope whatever of anything but an

accidental and transitory alleviation of any of our world's troubles. As mankind is, so it will remain until it puts its mind together. Our species may yet end its strange eventful history as just the last, the cleverest of the great apes. The great ape that was clever but not clever enough. It could escape from most things but not from its own mental confusion."

APPENDIX C

LIST OF PARTICIPANTS

ATHERTON, Pauline, Professor
School of Information Studies
Syracuse University
Syracuse, NY

BATTY, David C., Professor
College of Library & Info. Services
University of Maryland
College Park, MD

BEARMAN, Toni Carbo, Dr.
Executive Director
National Federation of Abstracting and Indexing Services
Philadelphia, PA

BERGER, Mary C.
Cuadra Associates
Santa Monica, CA

BERMAN, Sanford
Head Cataloger
Hennepin County Library
Edina, MN

BERNER, Richard C.
Head, University Archives & Manuscript Division
Univerity of Washington
Seattle, WN

BISHOFF, Lizbeth
Administrative Librarian
Ela Area Public Library District
Lake Zurich, IL

BOURNE, Charles, Dr.
Lockheed Information Systems
Palo Alto, CA

BRENNER, Everett H.
American Petroleum Institute
New York, NY

DURKIN, Kay
Manager, Marketing and User Services
Bibliographic Retrieval Services, Inc.
Ardmore, PA

EKSTROM, Ann
Director, Library Systems Division
OCLC, Inc.
Columbus, OH

EWING, Richard T.
Vice President - Editorial Director
Congressional Information Service
Washington, DC

FISCHER, Margaret T.
President, Management Decisions, Inc.
Greenwich, CT

GARFIELD, Eugene, Dr.
President
Institute for Scientific Information
Philadelphia, PA

GORDON, Lucille, Manager
Library Marketing & Promotion
McGraw-Hill Book Company
New York, NY

HENDRIKSON, Kent
Vice President, Corporate Operations
Blackwell/North America
Beaverton, OR

KRUPENIE, Paul H.
Physicist
National Bureau of Standards
Washington, DC

LAGUEUX, Paul B.
Information Systems Specialist
Council on Library Resources, Inc.
Washington, DC

LANCASTER, F. Wilfrid, Professor
Grad. School, Dept. of Library Science
University of Illinois
Urbana, IL

LINEBACK, Richard H., Dr.
Dir., Philosophy Documentation Cntr.
Bowling Green State University
Bowling Green, OH

LINFORD, John
Director, NELINET
Wellesley, MA

LUCKER, Jay K.
Director, The Libraries
Massachusetts Institute of Technology
Cambridge, MA

MESEROLE, Harrison, T., Dr.
Professor of English, Director,
Center for Research in Bibliographical and Textual Studies
Pennsylvania State University
University Park, PA

PAPENFUSE, Edward C.
Archivist
Hall of Records
Annapolis, MD

PARK, Margaret K., Dr.
Manager, Information Services
Office of Computing Activities
University of Georgia
Athens, GA

PIETRIS, Mary K.
Chief, Subject Cataloging Division
Library of Congress
Washington, DC

PRESCHEL, Barbara M., Index
Supervisor, Arete Publishing Co.
Princeton, NJ

ROSENTHAL, Frederick J.
Asst. Chief, Library Services Division
Congressional Research Service
Library of Congress
Washington, DC

ROWLETT, Russell, J. Jr., Dr.
Editor, Chemical Abstracts Service
Columbus, OH

SALTON, Gerard, Dr., Chairman
Department of Computer Science
Cornell Univesity
Ithaca, NY

SARACEVIC, Tefko, Dr.
Department of Library Science
Case Western Reserve University
Cleveland, OH

THIRD, Bettie Jane
Editor, Bus. Per. Index
The H. W. Wilson Company
Bronx, NY

VASTA, Bruno, Director
Chemical Information Division
Office of Toxic Substances
Environmental Protection Agency
Washington, DC

WIGINGTON, Ronald L., Dr.
Director, Research & Development
Chemical Abstracts Service
Columbus, OH

OBSERVERS

CHILD, Margaret, Dr.
Asst. Dir., Centers for Research
Programs, Div. of Research Grants
National Endowment for Humanities
Washington, DC

CIMA, Louis, Dr.
Associate Program Director
Research on Information Systems
National Science Foundation
Washington, DC

FARR, George F. Jr.
Asst. Dir., Centers for Research
Programs, Div. of Research Grants
National Endowment for Humanities
Washington, DC

GWINN, Nancy
Information and Publications Officer
Council on Library Resources, Inc.
Washington, DC

HAAS, Warren J.
President
Council on Library Resources, Inc.
Washington, DC

JONES, C. Lee
Program Officer
Council on Library Resources, Inc.
Washington, DC

RESNIKOFF, Howard L., Dr.
Director, Division of Information
Science & Technology
National Science Foundation
Washington, DC

ADMINISTRATIVE ASSISTANT
MC ILWAIN, Albert C., II
Council on Library Resources, Inc.
Washington, DC

REPORTER
RODERER, Nancy
King Research, Inc.
Rockville, MD